Rate Allocation for Cooperative Orthogonal-Division Channels with Dirty-Paper Coding

Cho Yiu Ng, *Student Member, IEEE*, Kenneth W. Shum, *Member, IEEE*, Chi Wan Sung, *Member, IEEE*, Tat Ming Lok, *Senior Member, IEEE*

***Abstract*—This paper investigates how much the rate region of the two-user Gaussian interference channel can be enlarged by allowing the two source nodes to cooperate. Two cooperative transmission schemes are proposed, based on dirty-paper coding and the assumption that the radio bandwidth is partitioned into two parts, and each part is utilized by one source node. The achievable rate regions and the outage performance of these two schemes are compared with the simplified Han-Kobayashi scheme, which is an efficient coding scheme for the interference channel. Simulation results show that in some channel realizations, the rate region of the Han-Kobayashi scheme is a subset of the rate regions of our two proposed cooperative transmission schemes. Furthermore, a significant gain in outage performance can be obtained, as the cooperative schemes have twice the diversity order of the simplified Han-Kobayashi scheme. While both cooperative schemes are able to yield large diversity gain, one of them can be implemented by simple decoder. Besides, it has an efficient algorithm for maximizing its weighted sum rate, and can be extended easily to the multi-channel case.**

***Index Terms*—Cooperative transmission, spatial diversity, achievable rate region.**

I. Introduction

COOPERATIVE transmission is a technique to increase the capacity of a wireless network. Briefly speaking, a number of wireless nodes form a coalition in which they exchange and cooperatively transmit messages. This concept stems from early information-theoretic works such as [1]. Recently, this idea was applied to the fading medium access channel (MAC) by Sendonaris *et al.* in [2]. They dealt with the case where two source nodes transmit their messages cooperatively to a common destination node. Their results showed that the sum rate can be significantly increased when users overhear the signals of others and help forward their data. Some protocols that exploit this phenomenon were proposed and analyzed in [3], [4].

Paper approved by E. Erkip, the Editor for Cooperation Diversity of the IEEE Communications Society. Manuscript received December 11, 2007; revised December 28, 2008 and January 10, 2010.

C. Y. Ng, K. W. Shum, and T. M. Lok are with The Chinese University of Hong Kong (e-mail: michaelng@ieee.org, kshum2010@gmail.com, tmlok@ie.cuhk.edu.hk).

C. W. Sung is with City University of Hong Kong (e-mail: itcwsung@cityu.edu.hk).

This work was supported in part by the grants from the Research Grant Council of the Hong Kong Special Administrative Region, China (Project no. CityU 120107 and CUHK 414807). This paper was presented in part at IEEE Global Telecommunications Conference 2007, Washington D.C., November 2007.

Digital Object Identifier 10.1109/TCOMM.2010.083110.070639

The most primitive form of cooperative transmission can be found in the relay channel, in which there are three nodes, one of them is a relay node, which helps the transmitter node forward data to the receiver node [5], [6]. *Full duplex* relay nodes are commonly assumed. That is, they can transmit and receive at the same time and frequency (e.g. [7], [8]). This assumption may be too strong for some applications, since it requires precise and expensive components at the relay node. This motivates the *half-duplex* model [9]–[11], which assumes that the relay node operates in time-division manner. For a given time period, the relay node is in the receive mode for a fraction of time and in the transmit mode for the remaining time. Alternatively, the relay node may operate in frequency-division manner in the sense that different portion of the frequency spectrum is used for transmit and receive modes [12], [13].

In this paper, we consider a more general channel model, which consists of two source nodes and two destination nodes. While full-duplex user cooperations in this setting are studied in [8], [14], we consider half-duplex transmitting nodes. In our model, each source node transmits in a set of channels orthogonal to another source node's channels. In addition, the source nodes overhear each other's channels for relaying the messages.

Our channel model encompasses some important three-node models as special cases. For example, it reduces to the MAC when the two destination nodes are physically collocated and regarded as one single node. User cooperation for this cooperative MAC is studied in [15]–[17]. Another example is this: When one of the source nodes does not have any messages to transmit, our model is reduced to a relay channel where the source and the relay nodes transmit on disjoint sets of orthogonal channels. The 95% outage capacity of this orthogonal relay channel is studied in [13]. Our work differs from [13] in that we assume channel state information can be estimated and passed among the nodes. Based on this information, we study how the source nodes allocate rate and power to individual link so as to support a given end-to-end data rate.

Our goal is to study whether performance gain can be obtained over the classical two-user interference channel by allowing the two source nodes to cooperate. Note that the capacity region of the interference channel, even for the two-user case, has long been an open problem. The largest rate region currently known is achieved with superposition coding and interference cancelation [18]. Etkin *et. al.* proposed a simplified Han-Kobayashi scheme in [19]. They have proved that the boundary of their achievable rate region is at most 1 bit/Hz from the capacity region. The authors also show that in many cases, this scheme can achieve the interference channel capacity region. In this paper, we will propose two new cooperative strategies for this channel and investigate the performance gain. Simulation results show that in some channel realizations, the achievable rate region of both proposed cooperative transmission schemes are larger than the interference channel capacity region. Moreover, it is also shown that the diversity order of one of the proposed schemes is twice of the one of the interference channel coding scheme proposed in [19].

The paper is organized as follows. We define the system model in Section II. The proposed transmission protocols are described in Section III. In Section IV, we compare the performance of the proposed transmission protocols with other schemes through simulations. After that, we propose a weighted sum rate maximization algorithm for one of the proposed transmission protocol in Section V. In the same section, we also describe how to extend that transmission protocol to the parallel channel case. Conclusions are then drawn in the last section.

II. System Model and Notations

We consider a wireless network with two transmitter-receiver pairs, denoted by (S_1, D_1) and (S_2, D_2). Node S_1 wants to transmit data to D_1 and node S_2 to D_2. We assume that the transmissions of S_1 and S_2 are on two orthogonal channels, each of bandwidth $B_W/2$ Hz[1], so that the total bandwidth is B_W Hz. Such orthogonality allows the source nodes to transmit their signals and overhear each other's signals simultaneously.

For $i = 1, 2$, consider the transmission of S_i. In the following, we let $j = 3 - i$, so that $\{i, j\} = \{1, 2\}$. For example, if $i = 2$, then $j = 1$. The power gain of the link from S_i to S_j, D_i, and D_j by a_i, b_i, and c_i, respectively. We will call a_i, b_i and c_i the *cooperative*, *direct* and *cross* link gains, respectively. Let the two-sided power spectral density of white noise experienced at each receiver be $N_0/2$ W/Hz. In a period of T seconds, each orthogonal channel has $B_W T$ real degrees of freedom [20, p.177], when $B_W T$ is large. In the channel where S_i is transmitting, the received channel symbols at S_j, D_i and D_j at time t, for $t = 1, 2, \ldots, B_W T$, are respectively

$$Y_{S_i S_j}[t] = \sqrt{a_i} X_{S_i}[t] + Z_{S_i S_j}[t] \quad (1)$$

$$Y_{S_i D_i}[t] = \sqrt{b_i} X_{S_i}[t] + Z_{S_i D_i}[t] \quad (2)$$

$$Y_{S_i D_j}[t] = \sqrt{c_i} X_{S_i}[t] + Z_{S_i D_j}[t], \quad (3)$$

where $X_{S_i}[t]$ is the transmitted symbol, and $Z_{S_i S_j}[t]$, $Z_{S_i D_i}[t]$ and $Z_{S_i D_j}[t]$ are additive white Gaussian noise (AWGN) with mean zero and variance $N_0/2$ (Fig. 1). Let P_i be the maximum

[1]For the convenience of presentation, we consider the equal bandwidth case. The results in this paper can be easily extended to the unequal bandwidth case.

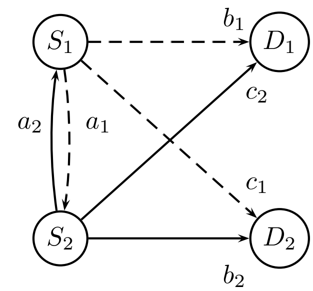

Fig. 1. System Model. The dashed lines correspond to the channel with S_1 as the sender, and the solid lines correspond to the channel with S_2 as the sender. The labels of the arrows are the associated link gains.

transmission power of S_i, for $i = 1, 2$, i.e., the transmitted signal satisfies

$$\frac{1}{B_W T} \sum_{t=1}^{B_W T} (X_{S_i}[t])^2 \leq P_i. \quad (4)$$

We call this model the *cooperative orthogonal-division channel*. The transmission rate from S_i to D_i is denoted by R_i, where $i = 1, 2$. In order to simplify notations, we normalize the power such that $N_0/2 = 1$, and assume that all noise powers are equal to 1.

We will use $\mathbb{R}_+^n$ to stand for the first orthant in the n-dimensional Euclidean space, i.e., it consists of the vectors with non-negative components. Let $C(x) \triangleq 0.5 \log_2(1 + x)$ denote the Shannon capacity formula. Vectors are typeset in boldface.

III. Transmission Schemes

We propose two transmission schemes in this section. The two schemes have the same encoding function, and differ in the the decoding processing. The encoding at the two source nodes are the same in both schemes. For the encoding, we note that the channel described by (1), (2) and (3) is a broadcast channel [6]. We will use an optimal coding scheme, called *nested lattice coding* [21], for the Gaussian broadcast channel as a building block. In the following, we first review some preliminaries on the Gaussian broadcast channel in general. After describing the common encoding process, we characterize the achievable rates by the two decoding methods.

A. Preliminaries on Gaussian Broadcast Channel

Consider a Gaussian broadcast channel (BC) with three users. For $i = 1, 2, 3$, the received symbol at destination node i is given by

$$y_i[t] = \sqrt{h_i} x[t] + z_i[t], \quad (5)$$

where $x[t]$ is the transmitted symbol at time t, $\sqrt{h_i}$ is the channel gain from the source node to node i, and $z_i[t]$ is an additive white Gaussian noise with zero-mean and unit-variance. The transmitted symbols over a block length of L symbols are subject to the power constraint

$$\frac{1}{L} \sum_{t=1}^{L} (x[t])^2 \leq P. \quad (6)$$

Let π be a permutation of $\{1,2,3\}$ such that

$$h_{\pi(1)} \geq h_{\pi(2)} \geq h_{\pi(3)}. \tag{7}$$

Suppose that the source node wants to send three independent messages to the three destination nodes, with rate r_1, r_2 and r_3. The capacity region of this Gaussian BC is

$$\bigcup \Big\{ (r_1, r_2, r_3) \in \mathbb{R}_+^3 : \\ r_{\pi(1)} \leq C(h_{\pi(1)}\alpha_{\pi(1)}P) \\ r_{\pi(2)} \leq C\Big(\frac{h_{\pi(2)}\alpha_{\pi(2)}P}{1 + h_{\pi(2)}\alpha_{\pi(1)}P}\Big) \\ r_{\pi(3)} \leq C\Big(\frac{h_{\pi(3)}\alpha_{\pi(3)}P}{1 + h_{\pi(3)}(\alpha_{\pi(1)} + \alpha_{\pi(2)})P}\Big)\Big\} \tag{8}$$

with the union taken over all non-negative real numbers α_1, α_2 and α_3 such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$ [6]. We denote this capacity region by $\mathcal{C}_{BC}(h_1, h_2, h_3, P)$.

The region defined by the union in (8) is a convex region. By the duality between Gaussian multiple-access channel and Gaussian BC [22], it can be shown that a rate pair (r_1, r_2, r_3) belongs to the capacity region in (8) if and only if it satisfies the following inequalities

$$0 \leq r_1 \leq C(h_1\beta_1 P) \tag{9}$$
$$0 \leq r_2 \leq C(h_2\beta_2 P) \tag{10}$$
$$0 \leq r_3 \leq C(h_3\beta_3 P) \tag{11}$$
$$r_1 + r_2 \leq C(h_1\beta_1 P + h_2\beta_2 P) \tag{12}$$
$$r_2 + r_3 \leq C(h_2\beta_2 P + h_3\beta_3 P) \tag{13}$$
$$r_1 + r_3 \leq C(h_1\beta_1 P + h_3\beta_3 P) \tag{14}$$
$$r_1 + r_2 + r_3 \leq C(h_1\beta_1 P + h_2\beta_2 P + h_3\beta_3 P) \tag{15}$$

for some non-negative real numbers β_1, β_2 and β_3 which sum to one.

There are several different encoding schemes which can achieve the above capacity region, e.g. superposition coding [23] and dirty-paper coding [24]. In superposition coding, the users are ranked according to their link gains. A user is required to first decode the message from the users with smaller link gains, subtract the corresponding signal from the received signal, and then decode his own message. In dirty-paper coding, a "pre-subtraction" is performed at the transmitter. The receivers are not required to know the message from the "weaker" users before decoding his own message bits. The decoders do not even need to know the codebook of the other users. This provides more flexibility in the design of the decoding method. Henceforth, we will focus on the dirty-paper coding for the Gaussian BC.

We outline in the following an efficient implementation of the dirty-paper coding, called nested lattice coding [21]. For the ease of discussion, we assume without loss of generality that $h_1 \geq h_2 \geq h_3$, so that user 1 has the best channel condition and user 3 is the worst user. The transmitter signal of length L is the sum of three signals

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3, \tag{16}$$

with the power of $\boldsymbol{x}_i$ equal to $\alpha_i P$, for $i = 1, 2, 3$, and non-negative real numbers α_1, α_2 and α_3 which sum to one. The encoding requires three pseudo-random "dither", denoted by U_1, U_2 and U_3, as inputs. For $i = 1, 2, 3$, the dither U_i is a shared randomness between the transmitter and user i, and is obtained by a pseudo-random generator. To encode three independent messages, say m_1, m_2 and m_3, we follow three steps. First, map m_3 to a codeword $\boldsymbol{x}_3(m_3, U_3)$, using the dither U_3 as one of the inputs. Secondly, interpret $x_3(m_3, U_3)$ as interference known non-causally, m_2 is dirty-paper coded on top of $x_3(m_3, U_3)$, and we obtain $\boldsymbol{x}_2(m_2, m_3, U_2, U_3)$. Finally, $\boldsymbol{x}_2(m_2, m_3, U_2, U_3)$ and $\boldsymbol{x}_3(m_3, U_3)$ are treated as non-causally known interference, and we apply dirty-paper coding again to m_1 and obtain the codeword $\boldsymbol{x}_1(m_1, m_2, m_3, U_1, U_2, U_3)$. As noted in [21, Section VI], although $\boldsymbol{x}_2$ is functionally dependent on $\boldsymbol{x}_3$, $\boldsymbol{x}_2$ is *statistically independent* of $\boldsymbol{x}_3$, due to the presence of the dither. Also, $\boldsymbol{x}_1$ is statistically independent of $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$.

As long as the rate triple (r_1, r_2, r_3) is within the capacity region (8), we can find a sufficiently large block length L, so that for $i = 1, 2, 3$, user i can decode m_i reliably with rate r_i. We refer the readers to [21] for more details.

B. Encoding at the Two Source Nodes

In our proposed transmission protocol, the encoding procedure at source nodes S_1 and S_2 are the same. In the following, we describe the encoding method for S_i, where i is equal to 1 or 2. The other source node is denoted by S_j, where $j \neq i$.

To encode its message, S_i splits its own data stream into two streams. The first stream is sent directly through the direct link between S_i and D_i. The second one is sent through a two-hop path, from S_i to the opposite source node S_j, and then from S_j to the intended destination D_i. Node S_j acts as a relay node, and re-encode the message to be forwarded. In other words, our proposed scheme can be classified as a *partial decode-and-forward* scheme; the relay node decodes and forwards only part of the message from the source node. We let r_{id} be the rate of data through the direct path, and r_{ir} the rate of data through the two-hop path. Here, the subscripts "$_d$" and "$_r$" signify "direct path" and "relay path" respectively.

Time is divided into $B+1$ time slots, each of which contains a codeword of length L. The data from S_i is divided into $2B$ parts: $b_{id}(n)$ and $b_{ir}(n)$, where $n = 1, 2, \ldots, B$. For each n, $b_{id}(n)$ consists of Lr_{id} bits, and is transmitted through the direct path from S_i to D_i; $b_{ir}(n)$ consists of Lr_{ir} bits, and is decoded and re-encoded by the opposite source node S_j.

From the viewpoint of S_i, it has to transmit three data streams: the first one is direct transmission to its intended receiver D_i; the second one is transmission to its relay node, S_j; the third one is forwarding the data from S_j to D_j. In time slot n, where $n = 1, 2, \ldots, B+1$, S_i transmits the codeword

$$\boldsymbol{x}_{S_i}(b_{ir}(n), b_{id}(n), \hat{b}_{jr}(n-1)), \tag{17}$$

where $\hat{b}_{jr}(n-1)$ denotes the decoded message $b_{jr}(n-1)$ from the previous time slot. Here, $\boldsymbol{x}_{S_i}$ is a codeword as in (16) from nested lattice coding, with three messages as inputs (the dithers are not shown for notational simplicity).

We initialize the encoding process by setting $b_{1r}(0)$ and $b_{2r}(0)$ to some known and constant bit strings, the all-zero bit strings for instance, of length Lr_{1r} and Lr_{2r} respectively. Likewise, the encoding process terminates in time slot $B+1$,

Slot 1	Slot 2	Slot 3	Slot 4
$b_{1r}(1)$	$b_{1r}(2)$	$b_{1r}(3)$	
$b_{1d}(1)$	$b_{1d}(2)$	$b_{1d}(3)$	
	$\hat{b}_{2r}(1)$	$\hat{b}_{2r}(2)$	$\hat{b}_{2r}(3)$

Slot 1	Slot 2	Slot 3	Slot 4
$b_{2r}(1)$	$b_{2r}(2)$	$b_{2r}(3)$	
$b_{2d}(1)$	$b_{2d}(2)$	$b_{2d}(3)$	
	$\hat{b}_{1r}(1)$	$\hat{b}_{1r}(2)$	$\hat{b}_{1r}(3)$

Fig. 2. Illustration of the encoding process ($B = 3$). The upper part indicates the messages encoded by S_1, and the lower part the messages by S_2.

by setting $b_{1r}(B+1)$, $b_{1d}(B+1)$, $b_{2r}(B+1)$, and $b_{2d}(B+1)$ to some pre-defined bit strings. An illustration of the encoding for $B = 3$ is shown in Fig. 2.

We note that there is a loss of data rate by a factor of $B/(B+1)$, which tends to 1 as B tends to infinity. Hence, this loss of data rate is negligible when B is large.

C. Achievable Rates

The rate region that can be achieved by our proposed encoding method depends on the decoding method. In this subsection, we will consider two decoding schemes. The first one is simpler but yields a smaller rate region. The second one is more complicated, as it requires joint decoding. In return, a larger rate region can be achieved.

In the first scheme, the received signals from the two orthogonal channels are decoded *separately*, using the decoding algorithm for nested lattice coding. The decoded data are put together to form the original message. In the sub-channel with S_1 as the source node, the rate triple (r_{1r}, r_{1d}, r_{2r}) is constrained by the capacity region of the corresponding Gaussian BC with link gains a_1, b_1 and c_1. Similarly, in the second sub-channel with S_2 as source node, the rate triple (r_{2r}, r_{2d}, r_{1r}) is limited by the Gaussian BC with link gains a_2, b_2 and c_2. We call this *Scheme 1* and characterize its achievable rates as follows.

Theorem 1 (Scheme 1): The rate pair $(R_1, R_2) \in \mathbb{R}_+^2$ is achievable by separately processing the received signals from the two orthogonal channels, if

$$R_1 = r_{1d} + r_{1r} \tag{18}$$

$$R_2 = r_{2d} + r_{2r} \tag{19}$$

$$(r_{1r}, r_{1d}, r_{2r}) \in 0.5 \cdot \mathcal{C}_{BC}(a_1, b_1, c_1, P_1) \tag{20}$$

$$(r_{2r}, r_{2d}, r_{1r}) \in 0.5 \cdot \mathcal{C}_{BC}(a_2, b_2, c_2, P_2). \tag{21}$$

(the product of a real number x and a set $\mathcal{S}$ is defined as $\{xy : y \in \mathcal{S}\}$.)

Proof: Equations (18) and (19) say that the total data rate is the sum of rates of the direct path and the relay path. The condition (20) and (21) mean that the rate vectors (r_{1r}, r_{1d}, R_{2r}) and (r_{2r}, r_{2d}, R_{1r}) are both feasible. The decoding and re-encoding at the two sources node is thus performed with arbitrarily small probability of error. We note that there is a factor of 0.5 in (20) and (21), because the total bandwidth is divided into two equal halves, one for each broadcast channel. ∎

In the second transmission scheme, the received signals from the two orthogonal channels are *jointly* processed in the decoding of the relayed message. For $n = 1, 2, \ldots, B$, the message $b_{1r}(n)$ is encoded by $\boldsymbol{x}_{S_1}(b_{1r}(n), b_{1d}(n), \hat{b}_{2r}(n-1))$ in the n-th time slot of the first channel, and $\boldsymbol{x}_{S_2}(b_{2r}(n+1), b_{2d}(n+1), \hat{b}_{1r}(n))$ in the $(n+1)$-st time slot of the second channel. The decoding function takes these two received signals as inputs and estimates $b_{1r}(n)$. Likewise, the decoding of $b_{2r}(n)$ is based on the received signals $\boldsymbol{x}_{S_2}(b_{2r}(n), b_{2d}(n), \hat{b}_{1r}(n-1))$ and $\boldsymbol{x}_{S_1}(b_{1r}(n+1), b_{1d}(n+1), \hat{b}_{2r}(n))$. This transmission scheme with joint processing of the signals from the two channels is called *Scheme 2*.

Since transmitter cooperative transmission is primarily effective when the link between the two source nodes are good, we will only state the rate region by Scheme 2 for the case where $a_i > b_i$ and $a_i > c_i$, for $i = 1, 2$. Let $\mathbb{I}$ be the indicator function defined by

$$\mathbb{I}(z) = \begin{cases} 1 & \text{if } z \text{ is true} \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Theorem 2 (Scheme 2): Suppose $a_i > b_i$ and $a_i > c_i$ for $i = 1, 2$, and assume without loss of generality that $b_i \neq c_i$, for $i = 1, 2$. The rate pair $(R_1, R_2) \in \mathbb{R}_+^2$ is achievable by Scheme 2 if it satisfies (18), (19) and the following conditions

$$r_{1r} \le \frac{1}{2} C(a_1 \alpha_1 P_1) \tag{23}$$

$$r_{2r} \le \frac{1}{2} C(a_2 \beta_1 P_2) \tag{24}$$

$$r_{1d} \le \frac{1}{2} C\Big(\frac{b_1 \alpha_2 P_1}{1 + b_1 P_1(\alpha_1 + \mathbb{I}(c_1 > b_1)\alpha_3)}\Big) \tag{25}$$

$$r_{2d} \le \frac{1}{2} C\Big(\frac{b_2 \beta_2 P_2}{1 + b_2 P_2(\beta_1 + \mathbb{I}(c_2 > b_2)\beta_3)}\Big) \tag{26}$$

$$r_{1r} \le \frac{1}{2} C(b_1 \alpha_1 P_1) + \frac{1}{2} C\Big(\frac{c_2 \beta_3 P_2}{1 + c_2 P_2(\beta_1 + \mathbb{I}(b_2 > c_2)\beta_2)}\Big) \tag{27}$$

$$r_{2r} \le \frac{1}{2} C(b_2 \beta_1 P_2) + \frac{1}{2} C\Big(\frac{c_1 \alpha_3 P_1}{1 + c_1 P_1(\alpha_1 + \mathbb{I}(b_1 > c_1)\alpha_2)}\Big) \tag{28}$$

for some non-negative real numbers α_1, α_2, α_3, β_1, β_2, and β_3 such that $\alpha_1 + \alpha_2 + \alpha_3 = \beta_1 + \beta_2 + \beta_3 = 1$.

Proof: The first two conditions in (23) and (24) ensure that the source nodes are able to decode the relay message from the opposite source node, so that the re-encoding is error-free. The conditions in (25) and (26) guarantee that the direct part of the data, $b_{id}(n)$, for $i = 1, 2$ and $n = 1, 2, \ldots, B$, can be sent reliably to the destination through the direct link. The last two conditions are the rate constraints for the decoding of the relay part of the data, $b_{ir}(n)$.

The first terms in (27) is the signal-to-noise ratio (SNR) of the signal

$$b_1 \cdot \boldsymbol{x}_{S_1}\big(b_{1r}(n), b_{1d}(n), \hat{b}_{2r}(n-1)\big) + \boldsymbol{z}_{S_1 D_1}, \tag{29}$$

received by D_1 in the first channel, where $\boldsymbol{z}_{S_1 D_1}$ denotes the noise vector with each component independently distributed according to the standard normal distribution. In the second term in (27), the fraction represents the SNR of

$$c_2 \cdot \boldsymbol{x}_{S_2}\big(b_{2r}(n+1), b_{2d}(n+1), \hat{b}_{1r}(n)\big) + \boldsymbol{z}_{S_2 D_1}. \tag{30}$$

Because of the use of dither, the two signals $\boldsymbol{x}_{S_1}$ and $\boldsymbol{x}_{S_2}$ in (29) and (30) are statistically independent. By standard argument from information theory, we see that the data rate on the right hand side of (27) is achievable after maximizing the mutual information between input and output. Similar comment goes with (28). ∎

We note that if R_{1d} and R_{2d} are restricted to zero, then Scheme 2 is the same as the transmission scheme in [25].

In the first proposed scheme, the decoding algorithm for the Gaussian BC is employed, and the capacity region $\mathcal{C}_{BC}$ appears in the statement of Theorem 1. In the second proposed scheme, in which the signals from the two orthogonal channels are jointly processed, the resulting rate region is strictly better than the one in the first scheme. However, the decoding complexity also increases accordingly.

IV. Performance Comparison

A. Achievable Rate Region

We compare the rate region achieved by Scheme 1 and 2 with a cut-set outer bound and the capacity region of the interference channel.

From [9], we have the following cut-set outer bound for the achievable rates. The derivation is straightforward and omitted. We refer the readers to [9] for more details.

Proposition 3 (Outer Bound): A rate pair (R_1, R_2) is achievable in the cooperative orthogonal-division channel only if it satisfies

$$R_1 \le 0.5 \cdot C((a_1^2 + c_1^2)P_1) \quad (31)$$
$$R_1 \le 0.5 \cdot C(b_1^2 P_1) + 0.5C(c_2^2 P_2) \quad (32)$$
$$R_2 \le 0.5 \cdot C((a_2^2 + c_2^2)P_2) \quad (33)$$
$$R_2 \le 0.5 \cdot C(c_2^2 P_1) + 0.5C(b_2^2 P_2) \quad (34)$$
$$R_1 + R_2 \le 0.5 \cdot [C((b_1^2 + c_1^2)P_1) + C((b_2^2 + c_2^2)P_2)]. \quad (35)$$

In non-cooperative transmission scheme, the link between source nodes S_1 and S_2 is ignored, and the channel reduces to the Gaussian *interference channel* (GIC). The two source nodes can transmit simultaneously, and the received symbols at the two destination nodes are

$$Y_{D_1}[t] = \sqrt{b_1} X_{S_1}[t] + \sqrt{c_2} X_{S_2}[t] + Z_{D_1}[t] \quad (36)$$
$$Y_{D_2}[t] = \sqrt{b_2} X_{S_2}[t] + \sqrt{c_1} X_{S_1}[t] + Z_{D_2}[t]. \quad (37)$$

For fair comparison, we consider interference channel with the same bandwidth as in Scheme 1 and Scheme 2, i.e., the total bandwidth is B_W, and the same total power constraint. In a period of T seconds, the number of real degrees of freedom is $2B_W T$. To make the total power constraint the same as Scheme 1 and 2, the average power of S_i in each channel symbol is $P_i/2$, for $i = 1, 2$.

The problem of finding the capacity region for the GIC in general is currently open. However, the answer is known in some special cases, for instance, the capacity region under strong interference [26], and the optimal sum rate in the low-interference regime [27]–[29]. We will compare with the capacity region of the GIC in the strong interference case. Suppose that $P_1 = P_2$, and $c_i > b_i$ for $i = 1, 2$. A rate pair (R_1, R_2) is achievable in the Gaussian interference channel

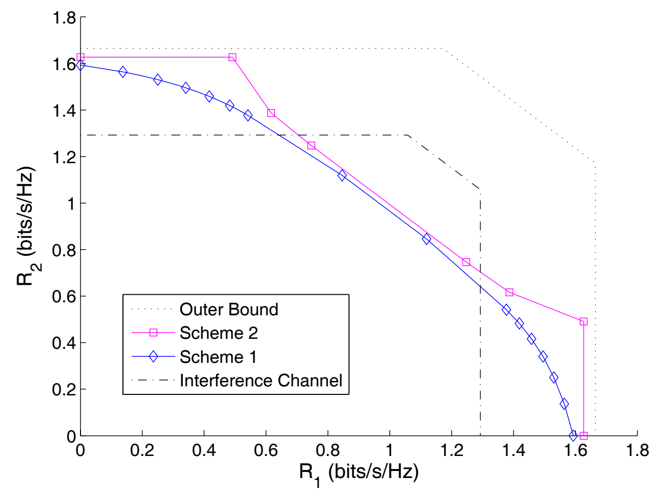

Fig. 3. Comparison of rate regions when the cooperative link is not very strong.

given in (36) and (37) if and only if

$$0 \le R_1 \le C(b_1 P_1/2) \quad (38)$$
$$0 \le R_2 \le C(b_2 P_2/2) \quad (39)$$
$$R_1 + R_2 \le \min\{C(\frac{b_1 P_1}{2} + \frac{c_1 P_2}{2}), C(\frac{b_2 P_2}{2} + \frac{c_1 P_1}{2})\}. \quad (40)$$

We remark that there we do not multiply the Shannon formulae in (38) to (40) by 0.5, because the the system occupies the whole bandwidth of B_W Hz.

In Fig. 3, we plot the cut-set outer bound, the rate regions of the two proposed schemes, and the capacity region of the GIC, for the case with parameters $P_1 = P_2 = 10$, and $a_i = 3$, $b_i = 1$ and $c_i = 2$ for $i = 1, 2$. The points for Scheme 1 are obtained by maximizing the weighted sum $w_1 R_1 + w_2 R_2$ over the feasible rate region, for different choices of weights w_1 and w_2. With the help of the duality theorem and the transformation for the capacity region Gaussian BC into (9)∼(15), the optimization can be performed by algorithm for standard convex programming. For scheme 2, we do not have efficient algorithm which maximizes weighted sum rate. A general maximization algorithm, called the *branch-and-bound* method, is applied instead.

We can see from Fig. 3 that the capacity region of the non-cooperative GIC and the rate regions of Schemes 1 and 2 do not dominate each other; the GIC have larger sum rate, whereas two cooperative schemes can achieve better rate pairs which are more asymmetric. We can also observe that the region of Scheme 2 defined as in Theorem 2 is not convex in this setting. The corner points $(0.4912, 1.6269)$ and $(1.6269, 0.4912)$ in Scheme 2 are of special interest. The rate pair $(0.4912, 1.6269)$ is achieved by setting the power allocation vector $(\alpha_1, \alpha_2, \alpha_3)$ to $(0, 9/11, 2/11)$ and $(\beta_1, \beta_2, \beta_3)$ to $(1, 0, 0)$. In the first subchannel, S_1 uses $9/11$ of the total power for its own message to be sent through the direct link, and $2/11$ of the total power for forwarding S_2's through the link from S_1 to D_2. In the second subchannel, S_2 uses all of its power for the message to be relayed. There is no direct message from S_2. Time-sharing can be applied if

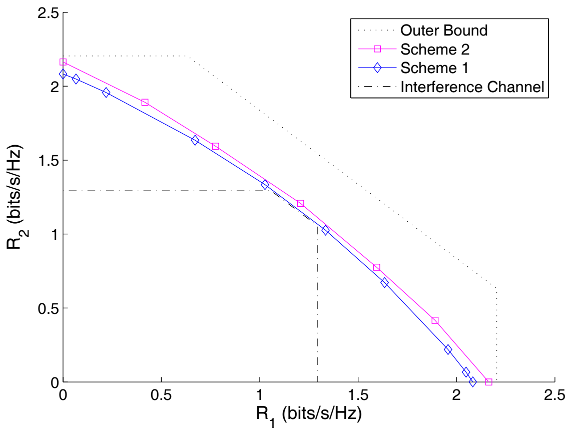

Fig. 4. Comparison of rate regions when the cooperative link is strong.

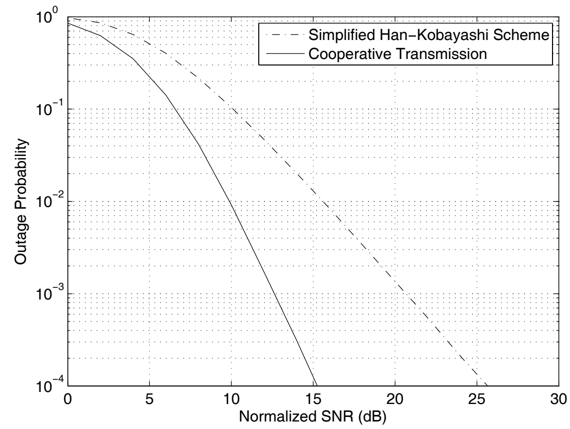

Fig. 5. Outage probability of Scheme 1 and the simplified Han-Kobayashi scheme ($R_T = 1$).

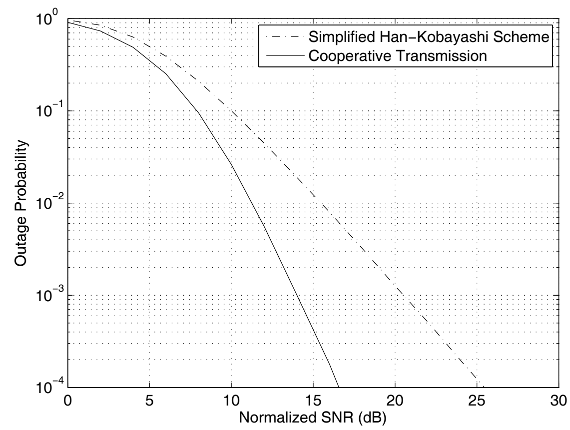

Fig. 6. Outage probability of Scheme 1 and the simplified Han-Kobayashi scheme ($R_T = 2$).

we want to operate on a point which lies on the line segment between $(0.4912, 1.6269)$ and $(1.6269, 0.4912)$.

In Fig. 4, we increase the cooperative link gain a_i from 3 to 10, and plot the corresponding rate region. We note that the capacity region for the GIC remains unchanged, because it does not depend on the link gains a_1 and a_2. However, the rate regions for Scheme 1 and Scheme 2 become larger, and include the capacity region of the GIC as a subset. It is in accordance with the heuristics that transmitter-cooperation is effective when the link between the two source nodes is good. We also observe that the gap between Scheme 1 and Scheme 2 is small.

B. Outage Performance

Now, we consider the outage performance. Due to the expensive computational cost for Scheme 2, we only compare Scheme 1 and a simplified Han-Kobayashi scheme proposed in [19] for interference channel through simulations. According to [19, Corollary 1], the achievable rate region contains all rate pairs (R_1, R_2) satisfying

$$R_i \le 2C\left(1 + \frac{b_i P_i}{2}\right) - 1, \quad i = 1, 2 \tag{41}$$

$$R_1 + R_2 \le \log_2\left(\frac{2c_1P_1 + b_1P_1}{2}\right) + 2C\left(\frac{2 + b_2P_2}{c_1P_1}\right) - 2 \tag{42}$$

$$R_1 + R_2 \le \log_2\left(\frac{2c_2P_2 + b_2P_2}{2}\right) + 2C\left(\frac{2 + b_1P_1}{c_2P_2}\right) - 2 \tag{43}$$

$$R_1 + R_2 \le 2C\left(\frac{c_2P_2}{2} + \frac{b_1}{c_1}\right) + 2C\left(\frac{c_1P_1}{2} + \frac{b_2}{c_2}\right) - 2 \tag{44}$$

$$2R_1 + R_2 \le 2C\left(\frac{b_1P_1 + c_2P_2}{2}\right) + 2C\left(\frac{c_1P_1}{2} + \frac{b_2}{c_2}\right) + 2C\left(1 + \frac{b_1}{c_1}\right) - 3 \tag{45}$$

$$2R_1 + R_2 \le 2C\left(\frac{b_2P_2 + c_1P_1}{2}\right) + 2C\left(\frac{c_2P_2}{2} + \frac{b_1}{c_1}\right) + 2C\left(1 + \frac{b_2}{c_2}\right) - 3 \tag{46}$$

The purpose is to compare the achievable order of diversity of these two schemes. Outage is defined as the event that the sum rate $R_1 + R_2$ is less than a given required sum rate R_T. In our simulations, R_T is chosen to be 1 and 2.

The power gains of all the links are independently and exponentially distributed with mean 1, which corresponds to the Rayleigh fading case. For simplicity, the transmission power of both source nodes are the same. We plot the outage probabilities of Scheme 1 and the simplified Han-Kobayashi scheme against the normalized SNR of each source node. If Γ is the common transmit SNR of the source nodes, the normalized SNR $\bar{\Gamma}$ is given by

$$\bar{\Gamma} = \frac{\Gamma}{2^{\frac{R_T}{2}} - 1}. \tag{47}$$

The denominator is the minimum required SNR of a source node at rate $\frac{R_T}{2}$ in the no fading case (i.e. an AWGN channel with power gain 1). The normalized SNR can be regarded as the additional amount of power (in dB) of each source node to combat against fading.

Let $\Phi_I(\bar{\Gamma})$ and $\Phi_C(\bar{\Gamma})$ be the outage probability of the simplified Han-Kobayashi scheme and Scheme 1 when the normalized transmit SNR is $\bar{\Gamma}$ respectively. We define the *diversity order* of Scheme 1 as follows. If there exists a real

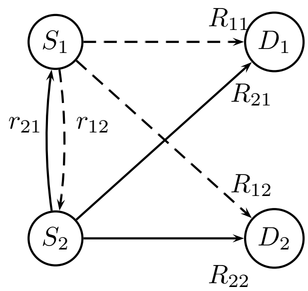

Fig. 7. Data rates in the System Model

number r such that

$$\lim_{\bar{\Gamma}\to\infty} \Phi_C(\bar{\Gamma})\bar{\Gamma}^k = r, \tag{48}$$

the diversity order of Scheme 1 is said to be k. If the diversity order of Scheme 1 is k, in the log-log plot of the outage probability against the normalized SNR, the slope of the curve of Scheme 1 is $-k$ for sufficiently large normalized SNR. The diversity order of the simplified Han-Kobayashi scheme is defined in a similar manner.

The results are plotted in Fig. 5 and 6. Both figures are obtained from 1,000,000 trials of Monte Carlo simulations. In both cases, the outage probability of Scheme 1 is strictly smaller than the simplified Han-Kobayashi scheme. Besides that, when the normalized SNR is greater than 15dB, the slope of the curves for Scheme 1 is roughly -4 while the slope of the curves for the simplified Han-Kobayashi scheme is -2. The diversity order of the simplified Han-Kobayashi scheme is 2 because there is a multi-user diversity among the two source-destination pairs. One reason why the diversity order of Scheme 1 is twice of the one of the simplified Han-Kobayashi scheme is that in an interference channel, the achievable rate of a source-destination pair is limited by the power gain of their direct link. If the direct link is in deep fading, there is no alternative path for the message to be transmitted. On the contrary, in Scheme 1, if the direct link is in deep fading, part of the messages can be relayed by another source node, which is a new path with independent fading. Therefore, the diversity order of Scheme 1 is twice of the one of the simplified Han-Kobayashi scheme.

V. WEIGHTED SUM RATE MAXIMIZATION FOR SCHEME 1

We have shown in the last section via numerical examples that the difference between Scheme 1 and Scheme 2 is small. In this section, we focus on Scheme 1, which can be implemented with smaller decoder complexity, and derive a fast weighted sum rate maximization algorithm.

A. An Iterative Algorithm Based on Lagrangian

To determine the optimal weighted sum rate for given weights w_1 and w_2, we propose an iterative algorithm based on Lagrangian duality. We slightly modify the rate constraints in Theorem 1 and formulate the weighted sum maximization as follows:

$$\text{maximize } w_1 R_1 + w_2 R_2 \tag{49}$$

subject to

$$R_1 = R_{11} + R_{21} \tag{50}$$
$$R_2 = R_{22} + R_{12} \tag{51}$$
$$(r_{12}, R_{11}, R_{12}) \in 0.5 \cdot \mathcal{C}_{BC}(a_1, b_1, c_1, P_1) \tag{52}$$
$$(r_{21}, R_{22}, R_{21}) \in 0.5 \cdot \mathcal{C}_{BC}(a_2, b_2, c_2, P_2) \tag{53}$$
$$r_{12} = R_{21} \tag{54}$$
$$r_{21} = R_{12}. \tag{55}$$

The notations are illustrated in Fig. 7.

It is readily checked that the constraints (50) to (55) are equivalent to those in Theorem 1.

Let $\boldsymbol{r} \equiv (r_{12}, R_{11}, R_{12}, r_{21}, R_{22}, R_{21})$ be a rate vector and $\boldsymbol{\mu} \equiv (\mu_1, \mu_2)$ be a vector of Lagrange multipliers. We relax constraints (54) and (55), and form the following *partial Lagrangian*:

$$\begin{aligned} L(\boldsymbol{r}, \boldsymbol{\mu}) \equiv\ & w_1(R_{11} + R_{21}) + w_2(R_{22} + R_{12}) \\ & + \mu_1(r_{12} - R_{21}) + \mu_2(r_{21} - R_{12}) \end{aligned} \tag{56}$$

for $\boldsymbol{r} \in \mathcal{R} \equiv \mathcal{C}_{BC}(a_1, b_1, c_1, P_1) \times \mathcal{C}_{BC}(a_2, b_2, c_2, P_2)$ and $\boldsymbol{\mu} \in \mathbb{R}^2$. We have the following *weak duality property*,

$$\max_{\boldsymbol{r}\in\mathcal{R}} \min_{\boldsymbol{\mu}\in\mathbb{R}^2} L(\boldsymbol{r}, \boldsymbol{\mu}) \le \min_{\boldsymbol{\mu}\in\mathbb{R}^2} \max_{\boldsymbol{r}\in\mathcal{R}} L(\boldsymbol{r}, \boldsymbol{\mu}), \tag{57}$$

which holds in general for any function of two sets of variables [30, p.379]. The max-min value on the left hand side of (57) is precisely equal to the optimal weighted sum in (49). It is because $\min_{\boldsymbol{\mu}\in\mathbb{R}^2} L(\boldsymbol{r}, \boldsymbol{\mu})$ is equal to $w_1(R_{11} + R_{21}) + w_2(R_{22} + R_{12})$ if the constraints $r_{12} = R_{21}$ and $r_{21} = R_{12}$ are satisfied, and $-\infty$ otherwise.

Consider the right hand side of (57). Let

$$q(\boldsymbol{\mu}) \equiv \max_{\boldsymbol{r}\in\mathcal{R}} L(\boldsymbol{r}, \boldsymbol{\mu}), \tag{58}$$

which is called the *dual function*. For each $\boldsymbol{\mu}$, the value of $q(\boldsymbol{\mu})$ is an upper bound of the maximum weighted sum rate.

By rearranging the terms, we can decompose the dual function $q(\boldsymbol{\mu})$ into $q(\boldsymbol{\mu}) = q_1(\boldsymbol{\mu}) + q_2(\boldsymbol{\mu})$ where

$$q_1(\boldsymbol{\mu}) \equiv \max_{r_{12}, R_{11}, R_{12}} \{\mu_1 r_{12} + w_1 R_{11} + (w_2 - \mu_2) R_{12}\} \tag{59}$$

$$q_2(\boldsymbol{\mu}) \equiv \max_{r_{21}, R_{22}, R_{21}} \{\mu_2 r_{21} + w_2 R_{22} + (w_1 - \mu_1) R_{21}\} \tag{60}$$

with the maxima taken over all $(r_{12}, R_{11}, R_{12}) \in \mathcal{C}_{BC}(a_1, b_1, c_1, P_1)$ and $(r_{21}, R_{22}, R_{21}) \in \mathcal{C}_{BC}(a_2, b_2, c_2, P_2)$ respectively. The computation of $q_1(\boldsymbol{\mu})$ and $q_2(\boldsymbol{\mu})$ amounts to the power allocation problems for maximizing the weighted sum rate in the broadcast channels with S_1 and S_2 as the sources respectively. Each of them can be solved by the greedy algorithm in [31], which is briefly described in Appendix A. Let $r^*_{ij}(\boldsymbol{\mu})$ and $R^*_{ij}(\boldsymbol{\mu})$ be the resultant solution. Hence, given any $\boldsymbol{\mu}$, $q(\boldsymbol{\mu})$ can be solved readily.

We will use the following fundamental theorem from the theory of Lagrangian multipliers [30, Theorem 28.3]. A short proof is included here for the sake of completeness.

Theorem 4 ([30]): If $\bar{\boldsymbol{\mu}}$ is a Lagrange multiplier vector such that

$$r_{12}^*(\bar{\boldsymbol{\mu}}) = R_{21}^*(\bar{\boldsymbol{\mu}}) \tag{61}$$

$$r_{21}^*(\bar{\boldsymbol{\mu}}) = R_{12}^*(\bar{\boldsymbol{\mu}}), \tag{62}$$

then the corresponding rate vector

$$\bar{\boldsymbol{r}} \equiv (\bar{r}_{12}, \bar{R}_{11}, \bar{R}_{12}, \bar{r}_{21}, \bar{R}_{22}, \bar{R}_{21}) \tag{63}$$

$$= (r_{12}^*(\bar{\boldsymbol{\mu}}), R_{11}^*(\bar{\boldsymbol{\mu}}), R_{12}^*(\bar{\boldsymbol{\mu}}), r_{21}^*(\bar{\boldsymbol{\mu}}), R_{22}^*(\bar{\boldsymbol{\mu}}), R_{21}^*(\bar{\boldsymbol{\mu}})) \tag{64}$$

which maximizes $L(\boldsymbol{r}, \bar{\boldsymbol{\mu}})$, is an optimal solution to the weighted sum rate maximization problem in (49) to (55).

Proof: From the weak duality property (57), it suffices to prove that

$$q(\bar{\boldsymbol{\mu}}) = L(\bar{\boldsymbol{r}}, \bar{\boldsymbol{\mu}}) \le \max_{\boldsymbol{r}\in\mathcal{R}} \min_{\boldsymbol{\mu}\in\mathbb{R}^2} L(\boldsymbol{r}, \boldsymbol{\mu}). \tag{65}$$

Putting (61) and (62) into the partial Lagrangian, we see that the equality

$$L(\bar{\boldsymbol{r}}, \boldsymbol{\mu}) = w_1(\bar{R}_{11} + \bar{R}_{21}) + w_2(\bar{R}_{22} + \bar{R}_{12}) = L(\bar{\boldsymbol{r}}, \bar{\boldsymbol{\mu}}) \tag{66}$$

holds for all $\boldsymbol{\mu} \in \mathbb{R}^2$. In particular, we get

$$L(\bar{\boldsymbol{r}}, \bar{\boldsymbol{\mu}}) = \min_{\boldsymbol{\mu}\in\mathbb{R}^2} L(\bar{\boldsymbol{r}}, \boldsymbol{\mu}), \tag{67}$$

which implies

$$L(\bar{\boldsymbol{r}}, \bar{\boldsymbol{\mu}}) \le \max_{\boldsymbol{r}\in\mathcal{R}} \min_{\boldsymbol{\mu}\in\mathbb{R}^2} L(\boldsymbol{r}, \boldsymbol{\mu}). \tag{68}$$

Thus, we have the following saddle-point property,

$$\max_{\boldsymbol{r}\in\mathcal{R}} \min_{\boldsymbol{\mu}\in\mathbb{R}^2} L(\boldsymbol{r}, \boldsymbol{\mu}) = L(\bar{\boldsymbol{r}}, \bar{\boldsymbol{\mu}}) = \min_{\boldsymbol{\mu}\in\mathbb{R}^2} \max_{\boldsymbol{r}\in\mathcal{R}} L(\boldsymbol{r}, \boldsymbol{\mu}). \tag{69}$$

This proves that the optimal weighted sum rate is achieved when $\boldsymbol{r} = \bar{\boldsymbol{r}}$, with maximal value $L(\bar{\boldsymbol{r}}, \bar{\boldsymbol{\mu}})$. ■

In view of Theorem 4, a vector of Lagrangian multipliers, $\boldsymbol{\mu}$, that satisfies the conditions in (61) and (62) is said to be *optimal*. Theorem 4 says that if $\bar{\boldsymbol{\mu}}$ is optimal, then the maximum weighted sum rate can be obtained from (66), and the associated optimal rate allocation is given by (64). In order to develop an iterative algorithm that computes the optimal $\boldsymbol{\mu}$, we first present the following lemma, which is about some continuous and monotonic properties of $r_{ij}^*(\boldsymbol{\mu})$ and $R_{ij}^*(\boldsymbol{\mu})$.

Lemma 5: Assume the noise power at the receivers of S_1, S_2, D_1 and D_2 are distinct. r_{12}^*, r_{21}^*, R_{12}^* and R_{21}^* are all continuous functions of $\boldsymbol{\mu}$. For $\mu_i \ge 0$, r_{ij}^* is an increasing function of μ_i and R_{ij}^* is a decreasing function of μ_i. For $\mu_i \le w_i$, r_{ji}^* is an increasing function of μ_i and R_{ji}^* is a decreasing function of μ_i.

Proof: See Appendix A. ■

The heuristic behind Lemma 5 is as follows. In the maximization of weighted sum rate in (59), the weighting of r_{12} is μ_1. If we increase μ_1, r_{12}^* will increase and R_{12}^* will decrease. The coefficient of R_{12} in (59) is $w_2 - \mu_2$. If we increase μ_2, the weighting of R_{12} is decreased. As a consequence, R_{12}^* will decrease but r_{12}^* will increase. Similar heuristic applies to (60).

Based on the above results, the optimal weighted sum is computed using an alternating optimization. We first fix μ_2 in the vector $\boldsymbol{\mu}$ and search for μ_1 such that (61) holds. Using the continuous and monotonic property in Lemma 5, this can be done by a simple binary search. We then fix the first component μ_1 in $\boldsymbol{\mu}$ and find μ_2 such that (62) holds. Again, by Lemma 5, a binary search suffices. We state the algorithm formally in Algorithm 1. The value of ϵ in Algorithm 1 is the error tolerance, and is set to a very small positive real number.

Algorithm 1 Search for Optimal $\boldsymbol{\mu}$

$t = 0$
$\mu_1(0) \leftarrow w_1$, $\mu_2(0) \leftarrow 0$
while $|r_{12}^*(\boldsymbol{\mu}(t)) - R_{21}^*(\boldsymbol{\mu}(t)| > \epsilon$ or $|r_{21}^*(\boldsymbol{\mu}(t)) - R_{12}^*(\boldsymbol{\mu}(t)| > \epsilon$ **do**
  search for ν_1 so that $r_{12}^*(\nu_1, \mu_2(t)) = R_{21}^*(\nu_1, \mu_2(t))$
  $t \leftarrow t + 1$
  $\mu_1(t) \leftarrow \nu_1$
  search for ν_2 so that $r_{21}^*(\mu_1(t), \nu_2) = R_{12}^*(\mu_1(t), \nu_2)$
  $\mu_2(t) \leftarrow \nu_2$
end while

Lemma 6: In each iteration t of Algorithm 1,

$$r_{12}^*(\boldsymbol{\mu}(t)) \ge R_{21}^*(\boldsymbol{\mu}(t)) \tag{70}$$

$$r_{21}^*(\boldsymbol{\mu}(t)) \le R_{12}^*(\boldsymbol{\mu}(t)). \tag{71}$$

In addition, $\mu_1(t)$ is decreasing with t while $\mu_2(t)$ is increasing with t.

Proof: See Appendix B. ■

Using the above lemmas, the convergence of the algorithm can then be proved.

Theorem 7: In Algorithm 1, $\mu_1(t)$ and $\mu_2(t)$ converge to the optimal solution.

Proof: We first show that $\boldsymbol{\mu}_1(t)$ and $\boldsymbol{\mu}_2(t)$ converges. For the convergence of $\boldsymbol{\mu}_1(t)$, we consider two cases. In the first case, we suppose that $\mu_1(t) < 0$ for some t, say at $t = t'$. We then have $r_{12}^*(\boldsymbol{\mu}(t')) = 0$. By (70) in Lemma 6,

$$0 = r_{12}^*(\boldsymbol{\mu}(t')) \ge R_{21}^*(\boldsymbol{\mu}(t')) \ge 0 \tag{72}$$

which means $r_{12}^*(\boldsymbol{\mu}(t')) = R_{21}^*(\boldsymbol{\mu}(t')) = 0$. This implies that $\mu_1(t) = \mu_1(t')$ for all $t \ge t'$. Hence $\mu_1(t)$ converges in this case. In the second case, suppose that $\mu_1(t) \ge 0$ for all t, i.e., $\mu_1(t)$ is lower bounded by 0. It has been shown in Lemma 6 that $\mu_1(t)$ is a decreasing function of t. Hence, if $\mu_1(t)$ is lower bounded by 0, it converges. Similarly, by considering the cases of $\mu_2(t)$ being or not being upper bounded by w_2, we can show that $\mu_2(t)$ is convergent.

After establishing the convergence of $\mu_1(t)$ and $\mu_2(t)$, we let the limit of $\mu_1(t)$ and $\mu_2(t)$ be $\bar{\mu}_1$ and $\bar{\mu}_2$ respectively. Because $r_{12}^*(\mu_1(t+1), \mu_2(t)) = R_{21}^*(\mu_1(t+1), \mu_2(t))$ for all $t \ge 0$ by construction, and r_{12}^* and R_{21}^* are continuous by Lemma 5, we obtain $r_{12}^*(\bar{\mu}_1, \bar{\mu}_2) = R_{21}^*(\bar{\mu}_1, \bar{\mu}_2)$. Therefore (61) holds with $\boldsymbol{\mu}_0 = (\bar{\mu}_1, \bar{\mu}_2)$. By the same argument and the fact that $r_{21}^*(\mu_1(t), \mu_2(t)) = R_{12}^*(\mu_1(t), \mu_2(t))$ for all $t \ge 1$, we can establish the equality in (62). The optimality of the solution then follows immediately from Theorem 4. ■

For positive w_1 and w_2, after running Algorithm 1, we obtain the limit of $\mu_1(t)$ and $\mu_2(t)$. Let $\bar{\mu}$ be the limit of $(\mu_1(t), \mu_2(t))$. We denote the optimal rate allocation which maximize $w_1 R_1 + w_2 R_2$ by

$$\check{R}_1(w_1, w_2) \equiv R_{11}^*(\bar{\boldsymbol{\mu}}) + R_{21}^*(\bar{\boldsymbol{\mu}}) \tag{73}$$

$$\check{R}_2(w_1, w_2) \equiv R_{22}^*(\bar{\boldsymbol{\mu}}) + R_{12}^*(\bar{\boldsymbol{\mu}}). \tag{74}$$

B. Extension to Parallel Channel Case

We generalize Scheme 1 to frequency selective channel, represented by a bank of parallel Gaussian channels. To be more specific, we consider the case that the source nodes employ multi-channel transmissions over disjoint and orthogonal frequency bands. Node S_1 has N_1 parallel sub-channels, whereas S_2 has N_2. If each parallel sub-channel represents one frequency carrier, then the model can be used to represent orthogonal frequency division multiplex (OFDM) transceivers. Let the bandwidth of each subchannel by B, so that the total bandwidth is equal to $B_W = B(N_1 + N_2)$.

Consider the transmission of S_i. Denote the power gain of subchannel k from S_i to S_j, D_i, and D_j by $a_i^{(k)}$, $b_i^{(k)}$, and $c_i^{(k)}$, respectively. Let P_i be the maximal transmission power of S_i, $i = 1, 2$. The total power P_i is split into N_i parts, $P_i^{(k)}$, $k = 1, 2, \ldots, N_i$, such that $P_i^{(k)}$ is the power associated with the k-th sub-channel of S_i, and

$$P_i^{(1)} + P_i^{(2)} + \ldots + P_i^{(N_i)} = P_i. \tag{75}$$

Each source node is associated with a parallel BC channel. The capacity region of the parallel BC channel for S_i is (76) with the union taken over all power allocations satisfying (75).

Similar to the single channel case, each source node S_i splits its own data stream into two streams: one direct to its intended receiver D_i and the other through a two-hop path with the other source node as relay. Each stream is divided into many blocks, each of which contains a codeword. The relay node decodes the received block from S_i, and then re-encodes and forwards a new block to D_i.

By a similar argument in Theorem 1, a rate pair

$$(R_1, R_2) = (r_{1r} + r_{1d}, r_{2r} + r_{2d}) \tag{77}$$

(with unit bits/s/Hz) is achievable if it satisfies

$$(r_{1r}, r_{1d}, r_{2r}) \in \frac{1}{N_1 + N_2} \mathcal{C}_1^{\|}(P_1) \tag{78}$$

$$(r_{2r}, r_{2d}, r_{1r}) \in \frac{1}{N_1 + N_2} \mathcal{C}_2^{\|}(P_2). \tag{79}$$

As in the single-channel case, we can solve the weighted sum rate maximization problem of this scheme by decomposing it into two weighted sum rate maximization problems for the two parallel Gaussian BCs. We can use the greedy algorithm in [31], which is applicable to the parallel Gaussian BCs.

VI. Conclusion

Cooperation between source nodes are explored for the Gaussian interference channel in this paper. We study two cooperative transmission schemes where the source nodes occupy disjoint bandwidth, and thus removing all interference. Simulation results show that in some channel realizations, the interference channel capacity region is a subset of the achievable rate regions of the two proposed schemes. Furthermore, one of the cooperative transmission schemes can achieve twice the order of diversity than a simplified Han-Koybayashi scheme. We end this manuscript by a generalization of this cooperative transmission scheme to the multi-channel case.

Appendix A
Proof of Lemma 5

We begin with a brief outline of the weighted sum rate maximization algorithm in [32]. Let $\mathcal{I}_i$ be the set of channels of user i. For $i = 1, 2$ and the $k \in \mathcal{I}_i$, let

$$u_{i1}^{(k)}(\boldsymbol{\mu}, z) = \frac{B^{(k)} \mu_i}{n_{i1}^{(k)} + z} - \lambda_i(\boldsymbol{\mu}), \tag{80}$$

$$u_{i2}^{(k)}(\boldsymbol{\mu}, z) = \frac{B^{(k)} w_i}{n_{i2}^{(k)} + z} - \lambda_i(\boldsymbol{\mu}), \tag{81}$$

$$u_{i3}^{(k)}(\boldsymbol{\mu}, z) = \frac{B^{(k)} (w_{3-i} - \mu_{3-i})}{n_{i3}^{(k)} + z} - \lambda_i(\boldsymbol{\mu}), \tag{82}$$

where $\lambda_i(\boldsymbol{\mu})$ is the solution of

$$\sum_{k \in \mathcal{I}_i} \left[\max \left\{ \left(\frac{B^{(k)} \mu_i}{\lambda_i(\boldsymbol{\mu})} - n_{i1}^{(k)} \right), \left(\frac{B^{(k)} w_i}{\lambda_i(\boldsymbol{\mu})} - n_{i2}^{(k)} \right), \left(\frac{B^{(k)} (w_{3-i} - \mu_{3-i})}{\lambda_i(\boldsymbol{\mu})} - n_{i3}^{(k)} \right) \right\} \right]^+ = P_i \tag{83}$$

with $[x]^+ = \max\{x, 0\}$. Here, each term in this summation is the optimal total amount of power allocated to all links in channel k.

Furthermore, for $i = 1, 2$ and $k \in \mathcal{I}_i$, let

$$u_i^{(k)*}(\boldsymbol{\mu}) = \left[\max_{1 \le j \le 3} \left\{ u_{ij}^{(k)}(\boldsymbol{\mu}, z) \right\} \right]^+, \tag{84}$$

$$\mathcal{A}_{ij}^{(k)}(\boldsymbol{\mu}) = \left\{ z \ge 0 | u_{ij}^{(k)}(\boldsymbol{\mu}, z) = u_i^{(k)*}(\boldsymbol{\mu}) \right\}, \quad 1 \le j \le 3. \tag{85}$$

The optimal r_{ij} and R_{ij}, $i \neq j$, for a given $\boldsymbol{\mu}$ are given by [31]

$$r_{ij}^*(\boldsymbol{\mu}) = \sum_{k \in \mathcal{I}_i} \int_{\mathcal{A}_{i1}^{(k)}(\boldsymbol{\mu})} \frac{B^{(k)}}{n_{i1}^{(k)} + z} dz, \tag{86}$$

$$R_{ij}^*(\boldsymbol{\mu}) = \sum_{k \in \mathcal{I}_i} \int_{\mathcal{A}_{i3}^{(k)}(\boldsymbol{\mu})} \frac{B^{(k)}}{n_{i3}^{(k)} + z} dz \tag{87}$$

where each term in the summations corresponds to the optimal rate allocated to that channel of the link. The optimal power allocated to the link from S_i to S_j in channel k is

$$\alpha_{i1}^{(k)} P_i^{(k)} = \int_{\mathcal{A}_{i1}^{(k)}} dz. \tag{88}$$

The optimal power for other links can be computed similarly.

As shown in [31], the set $\mathcal{A}_{ij}^{(k)}$ is an interval for all i, j, k. The continuity of r_{ij}^* and R_{ij}^* follows from (80)-(82), (84) and (85) together with the fact that an integral over an interval $[x, y]$ is a continuous function of x and y. This proves the first part of Lemma 5.

Let $\boldsymbol{\mu}^{(1)} = (\mu_1^{(1)}, \mu_2^{(1)})$ and $\boldsymbol{\mu}^{(2)} = (\mu_1^{(2)}, \mu_2^{(2)})$ such that $\mu_1^{(2)} = a\mu_1^{(1)}$, $a > 1$, $\mu_2^{(2)} = \mu_2^{(1)}$. Firstly, we would like to

$$\mathcal{C}_i^{\|}(P_i) \equiv \bigcup_{\sum_k P_i^{(k)}=P_i} \Big\{\boldsymbol{v}_1+\ldots+\boldsymbol{v}_{N_i} : \boldsymbol{v}_k \in \mathcal{C}_{BC}(a_i^{(k)}, b_i^{(k)}, c_i^{(k)}, P_i^{(k)})\Big\}, \tag{76}$$

show that $\lambda_1(\boldsymbol{\mu}^{(2)}) \le a\lambda_1(\boldsymbol{\mu}^{(1)})$. To begin with, we observe that for all $k \in \mathcal{I}_1$,

$$\frac{B^{(k)}\mu_1^{(2)}}{a\lambda_1(\boldsymbol{\mu}^{(1)})} - n_{11}^{(k)} = \frac{B^{(k)}\mu_1^{(1)}}{\lambda_1(\boldsymbol{\mu}^{(1)})} - n_{11}^{(k)} \tag{89}$$

$$\frac{B^{(k)}w_1}{a\lambda_1(\boldsymbol{\mu}^{(1)})} - n_{12}^{(k)} < \frac{B^{(k)}w_1}{\lambda_1(\boldsymbol{\mu}^{(1)})} - n_{12}^{(k)} \tag{90}$$

$$\frac{B^{(k)}(w_2-\mu_2^{(2)})}{a\lambda_1(\boldsymbol{\mu}^{(1)})} - n_{13}^{(k)} < \frac{B^{(k)}(w_2-\mu_2^{(1)})}{\lambda_1(\boldsymbol{\mu}^{(1)})} - n_{13}^{(k)}. \tag{91}$$

Hence, from (83),

$$\sum_{k\in\mathcal{I}_1}\left[\max\left\{\left(\frac{B^{(k)}\mu_1^{(2)}}{a\lambda_1(\boldsymbol{\mu}^{(1)})} - n_{11}^{(k)}\right), \left(\frac{B^{(k)}w_1}{a\lambda_1(\boldsymbol{\mu}^{(1)})} - n_{12}^{(k)}\right), \left(\frac{B^{(k)}(w_2-\mu_2^{(2)})}{2a\lambda_1(\boldsymbol{\mu}^{(1)})} - n_{13}^{(k)}\right)\right\}\right]^+ \le P_1. \tag{92}$$

Since the left hand side of the (83) is a monotonic decreasing function of λ_1 and $\lambda_1(\boldsymbol{\mu}^{(2)})$ must satisfy (83), $\lambda_1(\boldsymbol{\mu}^{(2)}) \le a\lambda_1(\boldsymbol{\mu}^{(1)})$. Similarly, we can also prove that $\lambda_1(\boldsymbol{\mu}^{(2)}) \ge \lambda_1(\boldsymbol{\mu}^{(1)})$. Thus, for all $k \in \mathcal{I}_1$ and z,

$$\frac{B^{(k)}\mu_1^{(2)}}{n_{11}^{(k)}+z} - \lambda_1(\boldsymbol{\mu}^{(2)}) = \frac{B^{(k)}a\mu_1^{(1)}}{n_{11}^{(k)}+z} - \lambda_1(\boldsymbol{\mu}^{(2)}) \tag{93}$$

$$\ge \frac{B^{(k)}a\mu_1^{(1)}}{n_{11}^{(k)}+z} - a\lambda_1(\boldsymbol{\mu}^{(1)}) \tag{94}$$

which implies that for all $k \in \mathcal{I}_1$ and z,

$$u_{11}^{(k)}(\boldsymbol{\mu}^{(2)}, z) \ge a u_{11}^{(k)}(\boldsymbol{\mu}^{(1)}, z). \tag{95}$$

As $\lambda_1(\boldsymbol{\mu}^{(2)}) \ge \lambda_1(\boldsymbol{\mu}^{(1)})$, it is trivial to see that for all $k \in \mathcal{I}_1$ and z,

$$u_{1j}^{(k)}(\boldsymbol{\mu}^{(2)}, z) \le u_{1j}^{(k)}(\boldsymbol{\mu}^{(1)}, z), \quad j = 2, 3. \tag{96}$$

Hence, for all $k \in \mathcal{I}_1$,

$$\mathcal{A}_{11}^{(k)}(\boldsymbol{\mu}^{(1)}) \subseteq \mathcal{A}_{11}^{(k)}(\boldsymbol{\mu}^{(2)}). \tag{97}$$

Since the right hand side of (86) is the sum of integrals of non-negative function,

$$r_{12}^*(\boldsymbol{\mu}^{(2)}) \ge r_{12}^*(\boldsymbol{\mu}^{(2)}). \tag{98}$$

Suppose $z' \in \mathcal{A}_{13}^{(k)}(\boldsymbol{\mu}^{(2)})$, $k \in \mathcal{I}_1$.

$$u_{13}^{(k)}(\boldsymbol{\mu}^{(2)}, z') \ge u_{12}^{(k)}(\boldsymbol{\mu}^{(2)}, z') \tag{99}$$

$$\frac{B^{(k)}(w_2-\mu_2^{(2)})}{n_{13}^{(k)}+z'} \ge \frac{B^{(k)}w_1}{n_{12}^{(k)}+z'} \tag{100}$$

$$\frac{B^{(k)}(w_2-\mu_2^{(1)})}{n_{13}^{(k)}+z'} - \lambda_1(\boldsymbol{\mu}^{(1)}) = \frac{B^{(k)}(w_2-\mu_2^{(2))}}{n_{13}^{(k)}+z'} - \lambda_1(\boldsymbol{\mu}^{(1)})$$

$$\ge \frac{B^{(k)}w_1}{n_{12}^{(k)}+z'} - \lambda_1(\boldsymbol{\mu}^{(1)}) \tag{101}$$

$$\therefore u_{13}^{(k)}(\boldsymbol{\mu}^{(1)}, z') \ge u_{12}^{(k)}(\boldsymbol{\mu}^{(1)}, z'). \tag{102}$$

Also, $u_{13}^{(k)}(\boldsymbol{\mu}^{(1)}, z') \ge u_{13}^{(k)}(\boldsymbol{\mu}^{(2)}, z') \ge u_{11}^{(k)}(\boldsymbol{\mu}^{(2)}, z') \ge u_{11}^{(k)}(\boldsymbol{\mu}^{(1)}, z')$. Therefore, $z' \in \mathcal{A}_{13}^{(k)}(\boldsymbol{\mu}^{(2)})$. This implies that

$$\mathcal{A}_{13}^{(k)}(\boldsymbol{\mu}^{(2)}) \subseteq \mathcal{A}_{13}^{(k)}(\boldsymbol{\mu}^{(1)}). \tag{103}$$

By the same argument as above,

$$R_{12}^*(\boldsymbol{\mu}^{(2)}) \le R_{12}^*(\boldsymbol{\mu}^{(1)}). \tag{104}$$

The relationship of r_{12}^*, R_{12}^* and μ_2 can be proved in the same manner as above.

Appendix B
Proof of Lemma 6

This can be proved by mathematical induction on t. Initially, $\mu_1 = w_1$ and $\mu_2 = 0$. This implies that $R_{21}^*(\boldsymbol{\mu}(0)) = 0$ as $w_1 - \mu_1 = 0$ and $r_{21}^*(\boldsymbol{\mu}(0)) = 0$. Hence,

$$r_{12}^*(\boldsymbol{\mu}(0)) \ge 0 = R_{21}^*(\boldsymbol{\mu}(0)) \tag{105}$$

$$R_{12}^*(\boldsymbol{\mu}(0)) \ge 0 = r_{21}^*(\boldsymbol{\mu}(0)) \tag{106}$$

which means it is true for $t = 0$.

Suppose it is true for $t \ge t'$ for some $t' \ge 0$, i.e.

$$r_{12}^*(\boldsymbol{\mu}(t)) \ge R_{21}^*(\boldsymbol{\mu}(t)), \tag{107}$$

$$r_{21}^*(\boldsymbol{\mu}(t)) \le R_{12}^*(\boldsymbol{\mu}(t)), \tag{108}$$

$\mu_1(t)$ is decreasing and $\mu_2(t)$ is increasing for $t \ge t'$.

Consider the iteration $t'+1$. Firstly, we update μ_1. According to the induction hypothesis, $r_{12}^*(\boldsymbol{\mu}(t')) \ge R_{21}^*(\boldsymbol{\mu}(t'))$. Therefore, we have to decrease μ_1 so that $r_{12}^*(\boldsymbol{\mu}(t'+1)) = R_{21}^*(\boldsymbol{\mu}(t'+1))$ according to Lemma 5.

According to Lemma 5 and the induction hypothesis, as μ_1 decreases,

$$r_{21}^*(\boldsymbol{\mu}(t'+1)) \le r_{21}^*(\boldsymbol{\mu}(t')) \le R_{12}^*(\boldsymbol{\mu}(t')) \le R_{12}^*(\boldsymbol{\mu}(t'+1)). \tag{109}$$

Therefore, after updating μ_1,

$$r_{12}^*(\boldsymbol{\mu}(t'+1)) = R_{21}^*(\boldsymbol{\mu}(t'+1)) \tag{110}$$

$$r_{21}^*(\boldsymbol{\mu}(t'+1)) \le R_{12}^*(\boldsymbol{\mu}(t'+1)) \tag{111}$$

$$\mu_1(t'+1) < \mu_1(t'). \tag{112}$$

Similarly, we can also prove that after updating μ_2,

$$r_{12}^*(\boldsymbol{\mu}(t'+1)) \ge R_{21}^*(\boldsymbol{\mu}(t'+1)) \tag{113}$$

$$r_{21}^*(\boldsymbol{\mu}(t'+1)) = R_{12}^*(\boldsymbol{\mu}(t'+1)) \tag{114}$$

$$\mu_2(t'+1) > \mu_2(t'). \tag{115}$$

Hence, the proposition is also true for $t = t'+1$. By mathematical induction, the proposition is true for all t.

References

[1] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 441-445, May 1983.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity–part I: system description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927-1938, Nov. 2003.

[3] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[4] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415-2425, Oct. 2003.

[5] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120-154, 1971.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[7] L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748-767, May 2004.

[8] A. Høst-Madsen, "Capacity bounds for cooperative diversity," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1522-1544, Apr. 2006.

[9] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "Bounds on achievable rates for general multi-terminal networks with practical constraints," in *Information Processing in Sensor Networks (IPSN '03)*, ser. Lecture Notes in Computer Sciences, F. Zhao and L. Guibas, editors, vol. 2634. Berlin Heidelberg: Springer-Verlag, Apr. 2004, pp. 146-161.

[10] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation in wireless relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020-2040, June 2005.

[11] I. Marić, R. D. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3536-3548, Oct. 2007.

[12] A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1815-1817, May 2005.

[13] A. Høst-Madsen, "On the capacity of wireless relaying," in *Proc. VTC 2002 Fall*, vol. 3, 2002, pp. 1333-1337.

[14] C. T. K. Ng and A. J. Goldsmith, "Transmitter cooperation in ad-hoc wireless networks: does dirty-paper coding beat relaying?" in *Proc. IEEE ITW 2004*, 2004, pp. 277-282.

[15] E. G. Larsson and B. R. Vojcic, "Cooperative transmit diversity based on superposition coding," *IEEE Commun. Lett.*, vol. 9, no. 9, pp. 778-780, Sep. 2005.

[16] K. Azarian, H. E. Gammal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4152-4170, Dec. 2005.

[17] O. Kaya and S. Ulukus, "Power control for fading cooperative multiple access channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2915-2923, Aug. 2007.

[18] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49-60, Jan. 1981.

[19] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534-5562, Dec. 2008.

[20] R. G. Gallager, *Principles of Digital Communication*. New York: Cambridge University Press, 2008.

[21] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250-1276, June 2002.

[22] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2657-2668, May 2003.

[23] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 197-207, Mar. 1973.

[24] W. Yu and J. M. Cioffi, "Trellis precoding for the broadcast channel," in *Proc. IEEE GLOBECOM*, vol. 2, New York, Nov. 2001, pp. 1344-1348.

[25] N. Fawaz, D. Gesbert, and M. Debbah, "When network coding and dirty paper coding meet in a cooperative ad hoc network," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1862-1867, May 2008.

[26] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inf. Theory*, vol. 6, no. 27, pp. 786-688, Nov. 1981.

[27] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 2, no. 55, pp. 689-699, Feb. 2009.

[28] A. S. Motahari and A. K. Khandani, "Capacity bounds for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 620-643, Feb. 2009.

[29] V. S. Annapureddy and V. V. Veeravalli, "Gaussian interference networks: sum capacity in the low-interference regime and new outer bound on the capacity region," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3032-3050, July 2009.

[30] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.

[31] D. N. C. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," in *IEEE Proc. Int. Symp. Inf. Theory*, Ulm, Germany, July 1997, p. 27.

[32] ——, "Optimal power allocation over parallel Gaussian broadcast channels." [Online]. Available: http://www.eecs.berkeley.edu/$\sim$dtse/broadcast2.pdf.

Cho Yiu Ng (S'08) received his B.Eng. and M.Phil. degrees in Information Engineering from The Chinese University of Hong Kong in 2003 and 2005 respectively. He was a research assistant in City University of Hong Kong and Chinese University of Hong Kong in 2006 fall and 2007 spring respectively. Since 2007 fall, he has been working towards the Ph.D. degree in Chinese University of Hong Kong. His research interests include OFDMA, cooperative transmission and network coding.

Kenneth W. Shum (M'00) received the B.Eng. degree in Information Engineering from the Chinese University of Hong Kong in 1993, and the M.S. and Ph.D. degree in Electrical Engineering from University of Southern California in 1995 and 2000 respectively. He is now a post-doctoral fellow in the Chinese University of Hong Kong. His research interests include information theory and resource allocation in wireless networks.

Chi Wan Sung (M'98) received the B.Eng, M.Phil, and Ph.D. degrees in information engineering from the Chinese University of Hong Kong in 1993, 1995, and 1998, respectively. Afterwards, he became an Assistant Professor in the same university. He joined the faculty at City University of Hong Kong in 2000, and is now an Associate Professor with the Department of Electronic Engineering. His research interests include multiuser information theory, design and analysis of algorithms, and optimization of wireless networks.

Tat Ming Lok received the BSc degree in electronic engineering from the Chinese University of Hong Kong, and the MSEE degree and the PhD degree in electrical engineering from Purdue University. In 1996, he joined the Chinese University of Hong Kong, where he is currently an associate professor. His research interests include communication theory, signal processing for communications and wireless communication systems. He has served on the technical program committees of many international conferences. He was a co-chair of the Wireless Access Track of IEEE VTC04 Fall. He served as an associate editor for the IEEE Transactions on Vehicular Technology for six years.